

# 控制与决策

Control and Decision

基于正交信号修正与高效偏最小二乘的质量相关故障检测方法

孔祥玉, 罗家宇, 张琪, 曹泽豪

引用本文:

孔祥玉, 罗家宇, 张琪, 等. 基于正交信号修正与高效偏最小二乘的质量相关故障检测方法[J]. 控制与决策, 2020, 35(5): 1167–1174.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.0708>

## 您可能感兴趣的其他文章

Articles you may be interested in

[基于复杂网络的多电机同步控制系统故障诊断与切换容错研究](#)

Fault diagnosis and switching tolerance for multi-motor synchronous control based on complex dynamical network  
控制与决策. 2020, 35(4): 843–851 <https://doi.org/10.13195/j.kzyjc.2018.0733>

[基于奇异值分解的非均匀采样非线性系统的模糊模型辨识](#)

Identification of fuzzy model of non-uniformly sampled nonlinear systems based on singular value decomposition  
控制与决策. 2020, 35(3): 757–762 <https://doi.org/10.13195/j.kzyjc.2018.0683>

[分组故障检测证据融合多星故障隔离算法](#)

Multi-satellite faults isolation algorithm based on group fault detection and evidences fusion  
控制与决策. 2019, 34(6): 1265–1270 <https://doi.org/10.13195/j.kzyjc.2017.1570>

[面向原油总氢物性预测的数据扩增预处理方法](#)

Data pretreatment approach for crude oil hydrogen properties prediction  
控制与决策. 2018, 33(12): 2153–2160 <https://doi.org/10.13195/j.kzyjc.2017.0937>

[多变量系统的耦合梯度辨识算法与性能分析](#)

Coupled stochastic gradient algorithm and performance analysis for multivariable systems  
控制与决策. 2016, 31(8): 1487–1492 <https://doi.org/10.13195/j.kzyjc.2015.0571>

[非均匀多采样率非线性系统的模糊辨识](#)

Fuzzy identification for non-uniformly multi-rate sampled nonlinear systems  
控制与决策. 2015(9): 1646–1652 <https://doi.org/10.13195/j.kzyjc.2014.1023>

[基于ESO迭代学习算法的故障估计](#)

Fault estimation method based on ESO iteration learning algorithm  
控制与决策. 2015(3): 546–550 <https://doi.org/10.13195/j.kzyjc.2014.0022>

[一种应用于故障诊断中的高效推理算法](#)

An effective inference algorithm for fault diagnosis  
控制与决策. 2015, 30(11): 2033–2040 <https://doi.org/10.13195/j.kzyjc.2014.1050>

# 基于正交信号修正与高效偏最小二乘的 质量相关故障检测方法

孔祥玉<sup>†</sup>, 罗家宇, 张 琪, 曹泽豪

(火箭军工程大学 导弹工程学院, 西安 710025)

**摘要:** 偏最小二乘(PLS)是一种广泛应用于多变量统计过程监控中的有效算法, 高效偏最小二乘(EPLS)是近年提出的一种 PLS 改进算法, 在质量相关故障检测中具有良好的检测效果, 但当测试数据含有质量无关故障时, EPLS 算法的误报率偏高, 可能导致误报警, 对工业过程中的故障检测有较大影响. 为降低检测质量无关故障的误报率, 将 EPLS 结合 4 种正交信号修正(OSC)方法提出 4 种 OSC-EPLS 算法. 用质量无关故障样本建立 OSC 模型对在线监测数据进行预处理, 将处理后的信息用 EPLS 算法进行故障检测, 误报率明显降低. 最后结合田纳西-伊斯曼工业过程, 应用 OSC-EPLS、PLS、EPLS 算法进行故障检测, 分别比较误报率和有效报警率的大小, 体现所提出算法在故障检测中的优势.

**关键词:** 故障检测; 正交信号修正; 质量相关; 误报率; 偏最小二乘; 过程监控  
**中图分类号:** TP273      **文献标志码:** A

## Quality-related fault detection method based on orthogonal signal correction and efficient PLS

KONG Xiang-yu<sup>†</sup>, LUO Jia-yu, ZHANG Qi, CAO Ze-hao

(Department of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China)

**Abstract:** The partial least squares algorithm (PLS) is an effective method, which has been widely used in multivariate statistical processes. Efficient projection to latent structures (EPLS) is an improved PLS algorithm proposed in recent years and has a good detection effect in quality-related fault detection. However, when the test data has a quality-independent failure, the false alarm rate of the EPLS algorithm is high, which may lead to false alarms and has a great influence on fault detection in industrial processes. In order to reduce the false alarm rate of detection-independent faults, four OSC-EPLS algorithms are proposed in combination with four orthogonal signal corrections (OSC) and the EPLS. In this paper, we establish an OSC model with quality-independent fault samples to preprocess on-line inspection data, and use the EPLS algorithm for fault detection to reduce the false alarm rate. Finally, combined with Tennessee Eastman industrial process, four kinds of the OSC-EPLS, PLS, and EPLS algorithms are used for fault detection. The false alarm rate is compared with the effective alarm rate to reflect the advantages of the proposed algorithm in fault detection.

**Keywords:** fault detection; OSC; quality-related; false alarm rate; partial least squares; process monitoring

## 0 引言

在传统工业系统过程中, 多元统计过程监控<sup>[1]</sup>(MSPM)是一种可以检测、诊断故障或异常情况的数据驱动过程方法, 其依托的主要理论是以主成分分析<sup>[2]</sup>(PCA)、偏最小二乘<sup>[3]</sup>(PLS)、规范变量分析<sup>[4]</sup>(CVA)等为核心的投影降维方法. 在处理高维度、高度耦合数据时具有独特的优势, 因此 MSPM 常被用于建立在大数据集正常运行条件下的过程监控模型, 监测过程中故障或异常等情况. 与传统工业

系统相比, 工业物联网(ICPS)面临着更多故障来源的严重威胁, Yu 等<sup>[5]</sup>总结了 MSPM 在 ICPS 中的研究进展, 并开发了 Matlab 工具箱 DK-KIT<sup>[6]</sup>对关键性能指标进行预测和诊断, 评估故障对系统的影响.

PLS 算法是 MSPM 中一种新型的多元统计分析方法, 该方法将输入空间  $X$  分解为主元子空间即质量相关子空间和残差子空间即质量无关子空间. 传统的故障检测方法是利用  $T^2$  统计量和  $Q$  统计量分别监测主元子空间和残差子空间中存在的异常情况. 标准

收稿日期: 2018-05-24; 修回日期: 2019-01-14.

基金项目: 国家自然科学基金项目(61673387, 61374120, 61833016, 61573366).

责任编辑: 陈虹.

<sup>†</sup>通讯作者. E-mail: xiangyukong01@163.com.

PLS在工业方面已经取得广泛应用,但是在理论中仍然存在以下两个缺陷:1)对 $X$ 空间进行斜交分解,导致主元子空间中仍包含了与 $Y$ 正交的成分;2)传统PLS算法基于提取 $X$ 和 $Y$ 协方差最大的潜变量,并未按照 $X$ 方差的大小顺序来抽取主元,因此残差子空间中存在较大的变异信息,不能准确检测出系统的质量变化。

为克服传统PLS存在的问题,Zhou等提出全潜结构投影<sup>[7]</sup>(T-PLS)算法,该算法将不同类的信息投影到不同的子空间中,对各子空间进行全方位监控.T-PLS模型没有改变对输出 $Y$ 的预测能力,但故障监测的效果在原来的基础上取得很大进步.在故障检测的实际应用中,PLS模型通常采用非线性迭代算法<sup>[8]</sup>(NIPALS)求解,导致计算量非常大,并且由于算法本身是斜交分解,使得主元中含有对预测无关的部分.基于此,Yin等<sup>[9]</sup>提出了改进潜结构投影(M-PLS)模型,该算法采用SVD分解得到正交投影算子,将 $X$ 正交投影到两个子空间中,避免了大量的迭代过程。

进一步研究发现,M-PLS在求解过程中存在广义逆计算,可能直接导致 $X$ 与 $Y$ 中相关的信息丢失.针对这一问题,Peng等<sup>[10]</sup>提出高效潜结构投影(EPLS)算法,在M-PLS的基础上对 $X$ 的残差子空间进行PCA分解,分别对残差子空间中的主成分和噪声也进行监控,全面提高质量相关故障的可检测性。

EPLS算法注重的是产品的质量,在进行故障检测时,会对与质量故障相关的输入潜变量进行检测.发生质量相关的故障不但会导致最终生产的产品不合格,甚至会对人身安全造成极大的危害.当检测到与质量相关的故障后,可以及时采取措施减少不必要损失,也可以预报危险事故的发生.然而EPLS算法基于一个近似的假设对 $X$ 进行与质量相关的正交分解,这使得分解得到的质量相关子空间中依然含有少量与质量无关的信息.在过程监控时,一些与质量无关的输入变量的变化依然轻微影响质量相关统计量的变化,从而出现了质量无关故障导致质量相关统计量超出控制阈值.但是,质量无关故障并不会影响最终的质量,因此发生误报警,其指标称为误报率.误报警会导致生产中断,特别是对连续流水生产形式的工业过程生产造成重大损失,因此如何降低误报率是实际生产中具有极大应用价值的研究内容。

为降低误报率,Wang等<sup>[11-12]</sup>在M-PLS算法基础上分别应用正交信号修正SWosc算法<sup>[13]</sup>和OPLS算法<sup>[14]</sup>,提出了OSC-MPLS算法<sup>[11]</sup>和MOPLS算法<sup>[12]</sup>.由OSC模型参数去除输入数据中与输出正交的信息,对处理后的数据进行故障检测,误报率得到有效

降低.可见,为降低EPLS的误报率,可以考虑直接降低输入数据中与质量正交的信息,将去除后的信息运用EPLS算法进行分解构造统计量,有效降低质量相关子空间中的质量无关信息.通过OSC算法对输入数据去除无用信息,再采用EPLS算法对测试数据进行故障检测,可以在EPLS算法的基础上进一步有效降低质量无关故障误报率。

目前有多种OSC方法可以去除输入数据与输出数据中的正交信息,OSC-MPLS和MOPLS分别是针对其中一类OSC算法展开的研究.每种算法去正交的原理并不相同,去除的正交信息量也各不相同,因此每类OSC算法结合EPLS对故障进行检测的效果也存在差异.为了比较每种算法在TEP中故障检测的效果,有必要对每种OSC-EPLS算法进行研究.本文分别引入SWosc、OPLS、OPLS(PCA)、JSosc四种正交修正算法结合EPLS提出4种OSC-EPLS算法,在不影响检测质量相关故障的情况下,降低测试数据中含有质量无关故障时的误报率.在OSC-EPLS算法中,采用正常数据建立EPLS模型,投影到3个子空间,分别负责监测质量相关的子空间、质量无关的子空间以及噪声子空间.提出用历史数据中质量无关故障样本建立OSC模型的新思想,去除检测数据中与输出数据正交的部分.在3个子空间中对检测数据设计合适统计量进行故障检测,并通过仿真实验给出结论性结果.本文创新体现在以下几点:1)提出4类OSC-EPLS算法;2)提出采用质量无关故障数据建立OSC模型的思想;3)结合4类算法和质量无关故障数据建立OSC模型的思想提出一套完整的过程监控技术,并对4类OSC-EPLS算法在过程监控中的监测性能进行详细对比分析,验证了所提出方法的有效性。

## 1 基础预备知识

### 1.1 标准PLS和M-PLS

假设自变量 $X = [x_1, \dots, x_m]$ ,因变量 $Y = [y_1, \dots, y_l]$ .其中: $x_i \in R^n, y_i \in R^n, i = 1, 2, \dots, N$ .PLS模型参数通常由NIPALS求解,用 $X$ 和 $Y$ 进行回归建模,有

$$\begin{cases} X = \hat{X} + \tilde{X} = TP^T + \tilde{X}, \\ Y = \hat{Y} + \tilde{Y} = TQ^T + \tilde{Y}. \end{cases} \quad (1)$$

其中: $\hat{X}$ 为主元子空间; $\tilde{X}$ 为残差子空间; $\hat{Y}$ 为可预测子空间; $\tilde{Y}$ 为残差; $T = [t_1, \dots, t_A]$ 为得分矩阵; $Q = [q_1, \dots, q_A]$ 为 $Y$ 对 $T$ 的回归矩阵,通常称为负载矩阵; $P$ 为 $X$ 对 $T$ 回归的负载矩阵.由于 $T$ 无法从 $X$ 中直接计算得到,引入权重矩阵 $R = W(P^T W)^{-1}$ ,可得 $T = XR$ ,基于 $R$ 可得 $Y$ 对 $X$ 的回归矩阵 $B = RQ^T$ .

$\hat{Y}$  计算如下:

$$\hat{Y} = TQ^T = X R Q^T = X B. \quad (2)$$

在 NIPALS 算法迭代过程中, 使用  $Y$  的特征向量回归  $X$  的特征向量, 相应地使用  $X$  的特征向量回归  $Y$  的特征向量, 通过旋转特征向量逐步提高相关性最终达到稳定.

PLS 模型参数求解无论基于特征值分解、SVD 分解或是 NIPALS 算法都会存在大量计算过程. 同时, 由于 PLS 模型斜交分解会导致质量相关子空间中仍然存在与质量正交的信息, 降低了故障监测性能.

M-PLS 算法正是针对这一缺陷的改进, 其要求主元子空间中不包含与  $Y$  正交的分量, 对预测  $Y$  有充分贡献; 残差子空间对预测  $Y$  没有贡献, 给出  $Y$  分解

$$Y = X M + E_y = \hat{Y} + E_y, \quad (3)$$

$$\frac{1}{N} Y^T X = \frac{1}{N} M^T X^T X + \frac{1}{N} E_y^T X \approx M^T \frac{X^T X}{N}, \quad (4)$$

$X$  和  $Y$  的系数矩阵为  $M = (X^T X)^{-1} X^T Y$ , 由于矩阵  $X^T X$  在计算中存在非满秩情况, 令  $M = (X^T X)^\dagger X^T Y$ , 对  $M M^T$  进行 SVD 分解, 有

$$M M^T = [\hat{P}_M \quad \tilde{P}_M] \begin{bmatrix} \Lambda_M & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{P}_M^T \\ \tilde{P}_M^T \end{bmatrix}. \quad (5)$$

其中  $\hat{P}_M \in R^{m \times l}$ ,  $\tilde{P}_M \in R^{m \times (m-l)}$ ,  $\Lambda_M \in R^{l \times l}$ . 构造正交投影矩阵  $\Pi_M$  和  $\Pi_M^\perp$ , 有

$$\Pi_M = \hat{P}_M \hat{P}_M^T, \quad (6)$$

$$\Pi_M^\perp = \tilde{P}_M \tilde{P}_M^T. \quad (7)$$

将  $X$  沿投影矩阵进行投影, 得到两个正交子空间

$$\hat{X} = X \Pi_M, \quad \tilde{X} = X \Pi_M^\perp. \quad (8)$$

上述过程将  $Y$  分解为与  $X$  相关的  $\hat{Y}$  和不相关的  $E_y$ , 然后将  $X$  向  $\text{Span}\{M\}$  和  $\text{Span}\{M\}^\perp$  投影正交分解成完全预测  $Y$  的子空间  $\hat{X}$  和与  $Y$  无关的子空间  $\tilde{X}$ , 该操作避免了 PLS 中大量复杂的迭代过程, 减少了计算量.

### 1.2 正交信号修正(OSC)算法

在工业过程中, 通常更注重质量相关故障导致的系统报警, 而多变量输入数据  $X$  通常包含与输出  $Y$  无关的变化, 可能导致系统发生误报警. 所以去除  $X$  中与  $Y$  无关的信息可以显著降低故障检测中潜变量的数量, 建立更加精确的 PLS 模型.

本节分别对 6 种正交信号修正算法进行介绍. OSC 算法最早称为 SWosc<sup>[10]</sup> 算法, 该算法的目的是为了去除  $X$  中与  $Y$  正交的信息; Sjöblom 等<sup>[15]</sup> 在 SWosc 算法基础上进行修改, 提出 JSosc 算法, 应用于校准转移; Fearn<sup>[16]</sup> 提出一种新的 OSC 算法, 该算法避免了迭代过程, 降低计算量, 简称为 TFosc<sup>[17]</sup> 算法; Andersson 等<sup>[17]</sup> 提出直接正交化 (DO) 算法, 该算法无迭代过程并且确保了预处理过程中  $X$  与  $Y$  完全相关的信息不会被删除, 缺点是模型的预测能力会降低; 文献[11]介绍了偏最小二乘正交投影 (OPLS) 算法和主成分分析偏最小二乘正交投影 (OPLS-PCA) 算法, OPLS 算法基于 NIPALS 正交校正, OPLS-PCA 算法是在 OPLS 算法的基础上对  $x$  的残差进行 PCA 分解并将残差重新加入  $X$  的主元中. 文献[18]指出, JSosc 算法和 DO 算法的正交程度较差, 在检测中可能去除多余信息降低故障检测率, 不适用于本文. 因此, 不对这两类算法步骤进行详细介绍, 给出其余 4 类 OSC 算法, 具体步骤见表 1.

表 1 OSC 算法步骤

Step	SWosc	TFosc	OPLS	OPLS(PCA)
1	标准化 $X$ 和 $Y$	标准化 $X$ 和 $Y$	标准化 $X$ 和 $Y$	标准化 $X$ 和 $Y$
2	PCA $\Rightarrow t$ , 令 $t_\perp = t$	$M = I - \frac{X'y}{y'X X'y} y' X Z = X M$	PLS( $X, y$ ) $\Rightarrow w, p$	PLS( $X, y$ ) $\Rightarrow w, p$
3	$t_{\text{new}} = t - \frac{y't}{y'y} y$	$Z = X M = X - \frac{X(y'X)'}{(y'X)(y'X)'} (y'X)$	$w_\perp = p - \frac{w'p}{w'w} w$	$w_\perp = p - \frac{w'p}{w'w} w$
4	PLS( $X, t_{\text{new}}$ ) $\Rightarrow W, P, q$	PLS( $X, t_{\text{new}}$ ) $\Rightarrow W, P, q$	$t_\perp = X w_\perp$	$t_\perp = X w_\perp$
5	$w_\perp = W(P'W)^{-1} q$	$w_\perp = p$	$P_\perp = X't_\perp / t_\perp' t_\perp$	$P_\perp = X't_\perp / t_\perp' t_\perp$
6	$t_\perp = X w_\perp$	$t_\perp = X w_\perp$	$X_{\text{OSC}} = X - t_\perp p'_\perp$	$X_{\text{OSC}} = X - t_\perp p'_\perp$
7	返回 step3 直到 $t_\perp$ 稳定	$P_\perp = X't_\perp / t_\perp' t_\perp$	$P_\perp = [P_\perp \quad p_\perp]$ $W_\perp = [W_\perp \quad w_\perp]$	$X_{\text{or}} = t_\perp p'_\perp, \text{PLS}(X_{\text{or}})$
8	$P_\perp = X't_\perp / t_\perp' t_\perp$	$X_{\text{OSC}} = X - t_\perp p'_\perp$		$X_{\text{or}} = t_{\text{pca\_or}} p_{\text{pca\_or}}^T + E_{\text{pca\_or}}$
9	$X_{\text{OSC}} = X - h t_\perp p'_\perp$	$P_\perp = [P_\perp \quad p_\perp] \quad W_\perp = [W_\perp \quad w_\perp]$		$X_{\text{OSC}} = X - t_\perp p'_\perp + E_{\text{pca\_or}}$
10	$P_\perp = [P_\perp \quad p_\perp]$ $W_\perp = [W_\perp \quad w_\perp]$			$P_\perp = [P_\perp \quad p_\perp]$ $W_\perp = [W_\perp \quad w_\perp]$

## 2 基于四类 OSC-EPLS 模型过程监控技术

### 2.1 OSC-EPLS 模型及其统计指标

EPLS 是 M-PLS 的改进算法, Peng 等<sup>[10]</sup> 在对

M-PLS 进行研究后指出, 由于  $M M^T$  采用 SVD 分解, 计算中  $(X^T X)^{-1}$  可能导致  $X$  中与  $Y$  相关的信息丢失. M-PLS 模型分解的  $\tilde{X}$  可能含有与  $Y$  相关的部分,

其变化会影响质量变量,因此有必要对 $\tilde{X}$ 进行PCA分解,提高质量相关故障的检测,即

$$\tilde{X} = T_{x_r} P_{x_r}^T + \tilde{X}_r, \quad (9)$$

其中 $T_{x_r}$ 和 $P_{x_r}$ 分别为 $\tilde{X}$ 的得分矩阵和负载矩阵.通过上述简要分析,建模如下:

$$X = \hat{X} + T_{x_r} P_{x_r}^T + \tilde{X}_r; \quad (10)$$

$$Y = \hat{Y} + E_y. \quad (11)$$

在EPLS算法中,输入数据由于噪声、扰动等情况导致模型不能完全正交分解,质量相关子空间仍会存在质量无关信息,出现监测质量无关故障误报率过高的情况.因此考虑直接采用OSC算法降低输入数据中与质量正交的信息,将处理后的数据应用EPLS模型进行故障检测.基于OSC-EPLS的监测模型可以有效降低质量相关子空间中的质量无关信息,因此认为将OSC算法引入EPLS是一个有效可行的思想,在EPLS算法的基础上进一步降低误报率.

第1.2节已比较了各类OSC算法的性能,因此选取正交性较好的SWosc、TFosc、OPLS、OPLS(PCA)算法分别构建基于OPLS(PCA)-EPLS、TFosc-EPLS、OPLS-EPLS、SWosc-EPLS的过程监控模型.OSC-EPLS过程监控统计量构造如下:有一组新测试数据 $x$ ,可以被分解为3个部分 $\hat{x}$ 、 $\hat{x}_r$ 、 $\tilde{x}_r$ , $\hat{x} \in R^{l \times 1}$ , $\tilde{x}_r \in R^{(M-l-A) \times 1}$ , $\hat{x}_r \in R^{A \times 1}$ , $A$ 为残差空间主元个数,由累计方差贡献率(CPV)确定,有

$$t_{\hat{x}}^2 = \hat{P}_M^T x, \quad (12)$$

$$t_{\hat{x}_r} = P_{x_r}^T \tilde{x}. \quad (13)$$

用 $t_{\hat{x}}$ 和 $t_{\hat{x}_r}$ 构造 $T^2$ 统计量, $\tilde{X}_r$ 构造 $Q$ 统计量<sup>[19]</sup>,有

$$T_{\hat{x}}^2 = t_{\hat{x}}^T \left( \frac{P_M^T X^T X P_M}{N-1} \right)^{-1} t_{\hat{x}}, \quad (14)$$

$$T_{\hat{x}_r}^2 = t_{\hat{x}_r}^T \left( \frac{T_{x_r}^T T_{x_r}}{N-1} \right)^{-1} t_{\hat{x}_r}, \quad (15)$$

$$Q_x = \|\tilde{x}\|^2 = \tilde{x}^T (I - P_x P_x^T) \tilde{x}. \quad (16)$$

统计量的控制限基于 $\chi^2$ 分布进行计算,有

$$J_{th, T_{\hat{x}}^2} = \hat{g} \chi_{\alpha, \hat{h}}, \quad \hat{g} = \frac{\hat{S}}{2\hat{\mu}}, \quad \hat{h} = \frac{2\hat{\mu}^2}{\hat{S}}; \quad (17)$$

$$J_{th, T_{\hat{x}_r}^2} = \hat{g}_r \chi_{\alpha, \hat{h}_r}, \quad \hat{g}_r = \frac{\hat{S}_r}{2\hat{\mu}_r}, \quad \hat{h}_r = \frac{2\hat{\mu}_r^2}{\hat{S}_r}; \quad (18)$$

$$J_{th, T_{\tilde{x}}^2} = \tilde{g} \chi_{\alpha, \tilde{h}}, \quad \tilde{g} = \frac{\tilde{S}}{2\tilde{\mu}}, \quad \tilde{h} = \frac{2\tilde{\mu}^2}{\tilde{S}}. \quad (19)$$

其中: $\hat{\mu}$ 和 $\hat{S}$ 、 $\tilde{\mu}$ 和 $\tilde{S}$ 、 $\hat{\mu}_r$ 和 $\hat{S}_r$ 分别为正态过程下统计量 $T_{\hat{x}}^2$ 、 $T_{\hat{x}_r}^2$ 和 $Q_x$ 对应的均值和方差.本节采用文献[7]中误报率(FAR)、有效报警率(FDR)两个指标对故障检测效果进行评估,公式如下:

$$FDR = N_{nea} / N_{tfs}, \quad (20)$$

$$FAR = N_{nfa} / N_{tfs}. \quad (21)$$

其中: $N_{nea}$ 和 $N_{nfa}$ 分别为有效报警和错误报警的数目, $N_{tfs}$ 为故障样本的总数.

在工业过程中,良好的故障检测效果体现在以下两方面:1)当检测数据是质量无关故障时,质量相关的统计指标FAR低;2)当检测数据是质量相关故障时,质量相关的统计指标FDR高.

## 2.2 基于质量无关故障数据的OSC模型

在故障检测中,通常采用正常数据建立OSC模型,将经OSC模型处理后的测试数据应用到故障检测中.但是经实验发现,由正常数据构建的OSC模型难以有效滤除故障数据中质量无关的信息,误报率不能得到有效降低.为此,提出由质量无关故障数据构建OSC模型的新思想,将历史中的故障数据分别用表1中给出的SWosc、TFosc、OPLS、OPLS(PCA)算法训练出模型参数 $W_{\perp}$ 、 $P_{\perp}$ .测试数据分别经4种OSC模型去除正交信息后得到新的输入数据 $X_{osc}$ ,将 $X_{osc}$ 投影到由EPLS模型分解的3个子空间,分别构建统计量进行故障监测.经检验,质量无关故障的检测效果有明显改善,误报率显著降低.

## 2.3 OSC-EPLS过程监控技术

本节采用2.1节介绍的统计量进行过程监控,见式(14)~(16),相应控制限由式(17)~(19)确定.现给出过程监控具体步骤如下.

step 1: 对输入数据 $X$ 和 $Y$ 进行标准化.

step 2: 对 $X$ 和 $Y$ 进行PLS非线性迭代,得到 $X$ 与 $Y$ 的系数矩阵 $M$ .

step 3: 将 $MM^T$ 进行SVD分解,得到 $\hat{P}_M$ 和 $\tilde{P}_M$ , $\hat{X} = X \Pi_M$ , $\Pi_M^{\perp} = \tilde{P}_M \tilde{P}_M^T$ .

step 4: 对 $X$ 进行正交分解, $X = \hat{X} + \tilde{X}$ , $\hat{X} = X \Pi_M$ , $\Pi_M^{\perp} = \tilde{P}_M \tilde{P}_M^T$ .

step 5: 对 $\tilde{X}$ 进行PCA分解, $\tilde{X} = T_{x_r} P_{x_r}^T + \tilde{X}_r$ , $\hat{X}_r = T_{x_r} P_{x_r}^T$ .

step 6: 计算控制限 $J_{th, T_{\hat{x}}^2}$ , $J_{th, T_{\hat{x}_r}^2}$ , $J_{th, T_{\tilde{x}}^2}$ .

step 7: 用质量无关的故障数据分别构建表1中的4种OSC模型,保存模型参数 $W_{\perp}$ 、 $P_{\perp}$ .

step 8: 当有一组测试数据 $x_{new}$ 时,用step 7中OSC模型参数 $W_{\perp}$ 、 $P_{\perp}$ 对 $x_{new}$ 进行预处理.

step 9:  $t_{new} = x_{new} w_{\perp} / (w_{\perp}^T w_{\perp})$ .

step 10:  $t_{new} = [t_{new} \quad t_{new}]$ .

step 11:  $x_{new-osc} = x_{new} - t_{new} P_{\perp}^T$ .

step 12: 令 $x_{new} = x_{new-osc}$ ,重复step 8~step 10,直到满足设定的去正交次数.

step 13: 由式 (15)~(17) 计算统计量  $T_{\hat{x}_{new}}^2$ ,  $T_{\hat{x}_{new-r}}^2$  和  $Q_x$ .

step 14: 诊断统计: 若  $T_{\hat{x}_{new}}^2 > J_{th, T_x^2}$ , 则检测到质量相关故障; 若  $T_{\hat{x}_{new-r}}^2 > J_{th, T_{\hat{x}_r}^2}$ , 则检测到质量无关故障.

基于 OPLS-EPLS、SWosc-EPLS、TFosc-EPLS、OPLS(PCA)-EPLS 算法的过程监控技术利用 4 种不同 OSC 算法去正交的性质, 分别去除输入数据  $X$  中与质量  $Y$  无关的信息. 将预处理后的信息投影到 EPLS 算法分解的 3 个子空间中, 结合 EPLS 算法在过程监控全面、高效的监测效果, 分别构建统计量进行过程监控. 与 EPLS 算法相比, OPLS-EPLS、OPLS(PCA)-EPLS、SWosc-EPLS、TFosc-EPLS 对质量无关故障误报警情况均有所改善, 系统鲁棒性明显提升, 稳定性进一步增强.

### 3 仿真实验

应用田纳西-伊斯曼过程 (Tennessee Eastman process, TEP) 验证 OPLS-EPLS、SWosc-EPLS、TFosc-EPLS 和 OPLS(PCA)-EPLS 算法在故障监测中的有效性.

#### 3.1 TEP 实验参数初始化

TEP 是由伊斯曼化工公司开发的一个工业模拟器, 整个流程由 5 个操作单元组成, 包括化学反应器、冷凝器、压缩机、汽/液分离器和分离器. 如文献 [20] 所述, TE 过程共生成 22 个数据集, 用于过程监控和故障诊断, 包括 1 个正常数据和 21 个特定故障数据集, 数据集中包括 41 个测量变量 XMEAS(1-41) 和 12 个控制变量 XMV(1-12); 正常数据集包含 480 个样本, 故障数据集包含 960 个样本. 由文献 [7] 可知, 在所有的故障样本中, IDV(1, 2, 5-8, 10, 12, 13) 是质量相关故障数据, IDV(3, 4, 9, 11, 14, 15) 是质量无关故障数据.

实验参数初始化: 选取数据集中 22 个过程变量 XMEAS(1-22) 和 11 个操纵变量 XMV(1-11) 作为输入  $X$ , 选取过程变量 XMEAS(35) 作为质量变量  $Y$ . WSosc 算法中 PLS 主元个数  $A_r = 7$ ; 用故障数据 IDV(4) 分别建立 4 种 OSC 模型 OPLS(PCA)-EPLS、SWosc-EPLS、TFosc-EPLS、OPLS-EPLS, 算法迭代次数均取  $noc = 7$ .

实验分为 3 部分, 实验 1 比较基于正常数据和质量无关故障数据建立 OSC 模型应用 OSC-EPLS 算法故障的监测效果; 实验 2 对质量相关故障和质量无关故障分别采用 PLS、EPLS 及 4 种 OSC-EPLS 算法的故障监测效果进行比较; 实验 3 验证 OSC-EPLS 对误报警情况的有效改善并与 TPLS 作对比. 最后对各算法

进行具体分析并给出结论性总结.

#### 3.2 实验 1

本节将检验质量无关故障和正常数据建立 OSC 模型的差异, 并探讨基于质量无关故障的 OSC 模型去除质量无关故障正交信息的有效性和合理性. 分别采用质量无关故障 IDV(3, 4, 9, 11, 15) 建立 SWosc 模型, 对预处理后的数据应用 EPLS 算法检测质量无关故障, 将各模型故障检测误报率分别与 EPLS 作比较, FAR 见表 2.

表 2 SWosc-EPLS 质量无关故障误报率 FAR %

OSC 建模数据类型	IDV(3)	IDV(4)	IDV(9)	IDV(11)	IDV(15)
EPLS	13.60	10.98	7.62	10.23	10.49
正常数据	12.23	11.24	6.37	10.36	9.99
IDV(3)	<b>7.24</b>	<b>9.11</b>	<b>5.49</b>	<b>8.61</b>	<b>9.36</b>
IDV(4)	<b>5.87</b>	<b>5.49</b>	<b>5.24</b>	<b>7.74</b>	<b>4.99</b>
IDV(9)	<b>12.23</b>	11.28	6.42	11.23	<b>7.24</b>
IDV(11)	<b>10.23</b>	<b>8.74</b>	<b>5.74</b>	11.61	<b>9.11</b>
IDV(15)	<b>6.24</b>	<b>8.61</b>	<b>5.37</b>	<b>9.99</b>	<b>8.24</b>

如表 2 所示, 黑体数据表示故障数据建立 OSC 模型在 TEP 中比正常数据建立 OSC 模型误报率低的监测值. 其中用正常数据建立 OSC 模型在监测故障 IDV(3, 9, 15) 比 EPLS 的误报率低, 但是在监测故障 IDV(4, 11) 时的误报率比 EPLS 高, 说明采用正常数据建立 OSC 模型的监测模型无法有效去除正交部分信息. 采用故障 IDV(3, 4, 11, 15) 建立的 OSC-EPLS 模型比 EPLS 模型和正常数据建立的 OSC-EPLS 模型对各类故障监测的误报率均进一步降低; 采用故障 IDV(9) 建立 OSC 模型对故障 IDV(3, 15) 监测的误报率比正常数据建立 OSC 模型和 EPLS 低, 只在监测故障 IDV(4, 11) 时误报率偏高. 综合上述数据分析可以得出, 基于质量无关故障数据建立 OSC 模型去除质量无关故障中正交的信息的效果都普遍较好, 过程监控 FAR 得到有效降低. 在质量无关故障中, 基于 IDV(4) 建立 OSC 模型去除各类质量无关故障正交信息效果最佳, 误报率均得到大幅度降低, 因此后面实验中采用故障 IDV(4) 建立 OSC 模型.

可以看出, 基于质量无关故障建立 OSC 模型去除质量无关故障中与质量正交信息是有效的方法, 因为在监测质量无关故障时导致 FAR 产生的原因都类似, 比如噪声、扰动以及过程的不稳定性等普遍客观存在的原因; 其次每类故障的发生必然是过程变量异常导致, 在同一个环境中这些异常变化会影响其他过程变量相关的故障, 如进料口温度发生阶跃变化引发的故障 IDV(3), 若该温度变化较大则可能影响反应

器冷凝水入口温度从而间接影响IDV(4),同时也可能导致冷凝器冷凝水阀门出现粘滞影响IDV(15),即故障之间存在相关关系. IDV(9,11)是随机扰动故障类型,采用阶跃响应故障IDV(3,4,15)建立的OSC模型去扰动故障正交信息能力较好,而基于随机扰动故障建立的OSC模型去正交性能力较差,这是由于随机扰动与其余故障相关性较差导致. 因此采用阶跃故障IDV(3,4,15)降低其他故障误报率是合理有效的.

### 3.3 实验2

本节实验将PLS算法、EPLS算法、SWosc-EPLS、TFosc-EPLS、OPLS(PCA)-EPLS、OPLS-EPLS算法应用到TEP对检测数据进行故障检测,分别比较检测质量相关故障和质量无关故障时的FDR和FAR值. 检测结果见表3和表4.

表3 质量相关故障有效报警率FDR %

IDV	PLS	EPLS	SWosc-EPLS	TFosc-EPLS	Opls-EPLS	Opls(pca)-EPLS
1	<b>99.88</b>	90.51	84.77	<b>86.14</b>	70.91	20.01
2	<b>98.88</b>	88.26	92.38	<b>96.00</b>	94.63	66.54
5	47.50	99.87	<b>99.88</b>	99.75	99.75	99.75
6	<b>99.88</b>	<b>99.00</b>	98.63	98.38	98.75	98.75
7	<b>99.88</b>	41.07	78.53	50.81	84.02	<b>88.14</b>
8	98.62	67.79	72.28	70.79	<b>73.40</b>	73.03
10	<b>89.25</b>	45.95	<b>52.43</b>	46.32	44.44	46.44
12	<b>99.65</b>	84.02	84.39	<b>83.52</b>	82.27	85.27
13	<b>95.38</b>	<b>90.26</b>	90.14	84.77	84.77	85.89

表4 质量无关故障误报率FAR %

IDV	PLS	EPLS	SWosc-EPLS	TFosc-EPLS	Opls-EPLS	Opls(pca)-EPLS
3	9.50	13.60	<b>4.49</b>	6.99	5.12	6.62
4	100	10.98	5.62	6.99	<b>5.49</b>	9.98
9	7.25	7.62	5.36	5.12	<b>3.87</b>	6.99
11	82.38	10.23	<b>7.11</b>	9.11	7.11	9.61
15	17.00	10.49	5.12	6.99	<b>4.62</b>	8.74

表3黑体数据表示FDR最大值以及除PLS外的FDR最大值. 由表3可知:检测数据中含有质量相关故障时,PLS算法对IDV(1,2,6,7,8,10,12,13)均有较好的检测率;EPLS算法检测故障数据IDV(5,6)时有较好的报警效果,但是检测IDV(7,8,10,12)时FDR偏小,当测试数据中含有此类故障时,不能有效进行报警,因此EPLS算法对质量相关故障的监测能力仍有待提高;OPLS(PCA)-EPLS算法在检测故障数据IDV(1,2)时的FDR值与其余算法相比有效报警性能最差,当检测中含有该二类故障时可能导致漏报等情况;OPLS-EPLS算法在检测IDV(1)时报警率低,当故障发生时不能对其有效报警,而该算法在检测其余故障时均有较高的检测率,总体上对质量相关

故障有较好的检测性能;TFosc-EPLS对故障IDV(7)的检测率较低,对其余各类故障均有较好的监测效果;SWosc-EPLS在故障监测时,对每类质量相关故障均有较高检测率;当设定一定阈值判断系统是否发生故障时,PLS算法与SWosc-EPLS算法对各类质量相关故障均能有效报警.

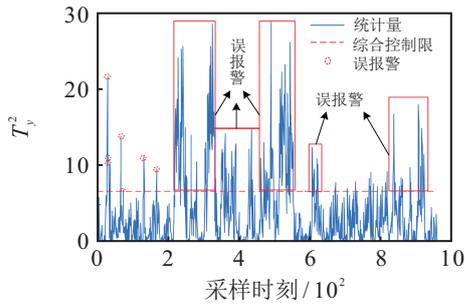
当测试数据中含有质量无关故障时不会影响质量的变化,所以较小的质量无关故障检测率代表了更为优异的监测性能,避免质量无关故障幅值较大导致系统发生误报警. 表4黑体数字表示检测数据为质量无关故障时的最低误报率,即出现误报警的情况较低. 由表4可知:PLS算法在检测质量无关故障IDV(4,11,15)时的FAR值极大,当测试数据中含有此类故障时易导致误报警;EPLS算法在检测故障IDV(4,11,15)时的FAR值均得到较大降低,误报警情况有一定改善,但是在检测故障IDV(3)时的误报率过大,测试数据含有此类故障时易出现误报警的情况;将OSC结合EPLS对各类质量无关故障进行检测与EPLS相比均有较大提升,其中SWosc-EPLS在IDV(3,11)时误报率最小,当发生故障IDV(3,11)时,系统鲁棒性强,不易发生误报;OPLS-EPLS检测IDV(4,9,15)的误报率最小,此类故障发生时该算法检测效果最为良好;TFosc-EPLS算法与OPLS(PCA)-EPLS的误报率与EPLS相比均有较大改善. 综上,当检测数据中含有质量无关故障时,SWosc-EPLS和OPLS-EPLS的误报率较低,有更为优异的检测效果.

### 3.4 实验3

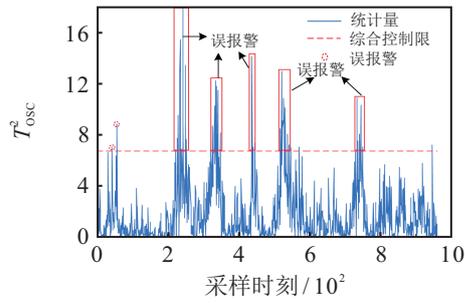
为检验OSC-EPLS对测试数据误报警的监测情况,将基于T-PLS算法在故障监测中全面优良的监测性能,与OSC-EPLS算法在质量无关故障中的检测效果进行对比分析. 应用TEP仿真,OSC采用SWosc模型,质量无关故障取IDV(4)和IDV(11),T-PLS采用文献[7]中具体参数,对质量相关子空间 $T_y^2$ 和 $T_{osc}^2$ 监测情况如图1和图2所示.

图1、图2分别为对故障IDV(4),IDV(11)的监测结果,其中前160个数据为正常数据,该部分数据发生误报警用红色虚线圆标出,后800个数据为质量无关故障,该部分数据超过控制限代表出现误报警,用红色实线方框圈出.

图1(a)中,T-PLS出现少量监测正常数据产生的误报警,而质量无关故障样本误报警严重,统计量大量超过控制限. 图1(b)中,SWosc-EPLS基本无正常数据导致的误报警,并且质量无关故障误报警情况与T-PLS相比有明显降低. 图2(a)中,T-

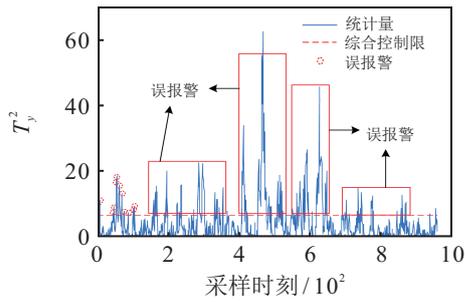


(a) T-PLS 故障检测  $T_y^2$  统计量

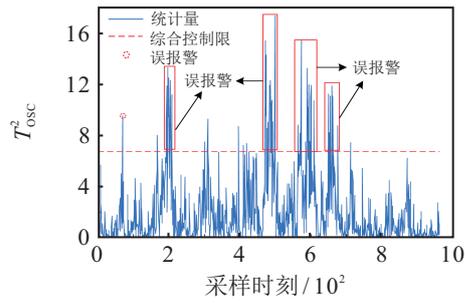


(b) SWosc-EPLS 故障检测  $T_{osc}^2$  统计量

图 1 故障 IDV(4) 监测图



(a) T-PLS 故障检测  $T_y^2$  统计量



(b) SWosc-EPLS 故障检测  $T_{osc}^2$  统计量

图 2 故障 IDV(11) 监测图

PLS 正常数据导致误报警样本数较多、波动较大,并且监测质量无关故障的误报警严重.图2(b)中,SWosc-EPLS 正常数据统计量总体均处于控制限下,误报警样本点只有1个并且质量无关故障的误报警有显著改善,红框中的误报警样本点数量与T-PLS相比数量大幅降低.综合图1、图2可以发现,误报警由正常数据或质量无关故障数据误报导致,而基于OSC-EPLS的故障监测方法均可有效改善这两类误报警.

### 3.5 算法结论性总结

本节综合实验1、实验2、实验3,对每种OSC-EPLS算法进行总结,并在最后给出结论性总结.

OPLS(PCA)-EPLS 算法在检测质量相关故障IDV(1,2)时有效报警率低,若测试数据中含有此类故障,则可能出现漏报的情况.在检测质量无关故障时,该算法的误报率相比EPLS算法有一定提高,但仍低于其余的OSC-EPLS算法,所以在实际应用中不是理想的算法.

TFosc-EPLS 算法在检测质量相关故障IDV(2,12)时有效报警率最大,且对其余故障有效报警率也较高,所以该算法对故障的报警是有效可取的.在检测质量无关故障时,各类故障的误报率较EPLS算法也有一定降低,但仍高于SWosc-EPLS算法和OPLS-EPLS算法,若系统报警阈值较高,则此算法对过程数据的检测将会取得良好的效果.

OPLS-EPLS算法在检测质量相关故障IDV(1)时有效报警率低,测试数据含有此类故障时可能存在误报警的情况,但是该算法对其余各类质量相关故障的检测效果均较为良好,仍是一个可行的检测故障的方法.在检测质量无关故障IDV(4,9,15)时误报率最低,去正交信息效果十分明显.综合来看,该算法可以有效应用到工业过程故障监测中.

SWosc-EPLS 算法在检测质量相关故障IDV(1,5,10)有效报警率最大,其余各类故障的报警率也较高,对比其余算法,该算法在质量相关故障中的检测效果最好.在检测质量无关故障IDV(1,11)误报率最小,其余质量无关故障误报率也较低.综合来看,SWosc-EPLS算法是一个非常有效的故障监测方法.

总体来看,采用OSC算法去除正交信息后应用到EPLS算法的故障检测中是一个非常有效的方法,对工业过程监测误报警情况有较大的改善.在每类OSC-EPLS算法中,OPLS(PCA)-EPLS算法去除正交信息效果较差,导致故障检测效果不佳,通常情况不适用于故障检测;OPLS-EPLS算法在质量相关故障的检测存在部分缺陷;TFosc-EPLS算法在质量无关故障检测中的误报率偏低,在条件阈值允许情况下,可以选择用于故障检测;SWosc-EPLS算法经仿真验证,无论是质量相关故障还是质量无关故障该算法都具有非常良好的检测效果,是较为理想的工业过程故障检测算法.

### 4 结论

本文将4种OSC算法分别与EPLS算法相结合,提出了基于SWosc-EPLS、TFosc-EPLS、OPLS-EPLS、

OPLS(PCA)-EPLS的质量相关故障监测方法. 首先利用质量无关故障数据建立4种OSC模型, 分别对测试数据进行预处理; 然后由正常数据建立EPLS模型, 构造合适统计量对TEP进行故障监控. 所提出的4种OSC-EPLS算法与T-PLS算法相比, 计算量较低, 误报率有效降低. 与EPLS算法相比, 不仅在其基础上进一步降低误报率, 而且质量相关故障的有效报警率也得到提高, 质量相关和无关故障的检测效果均得到改善, 基于4类OSC-EPLS算法的过程监控技术在工业应用中有着良好的检测效果. 其中, WSosc-EPLS算法在4类算法中有较为优异的监测效果, 可以应用到工业过程的故障监测中.

在本文的研究中, 对正交次数和OSC算法中主元个数的确定还是一个难题, 也是所提出算法有待提高的研究方向, 因此可以进一步通过PSO等优化算法来自适应确定最优参数, 从而得到更好的故障检测效果.

#### 参考文献(References)

- [1] 李晗, 萧德云. 基于数据驱动的故障诊断方法综述[J]. 控制与决策, 2011, 26(1): 1-9.  
(Li H, Xiao D Y. Survey on data driven fault diagnosis methods[J]. Control and Decision, 2011, 26(1): 1-9.)
- [2] Jolliffe I T. Principal component analysis and factor analysis[M]. New York: Springer, 1986: 115-128.
- [3] Wold S, Martens H, Wold H. The multivariate calibration problem in chemistry solved by the PLS method[M]. Berlin Heidelberg: Springer, 1983: 286-293.
- [4] Ms L H C, Russell E L, Braatz R D. Fault detection and diagnosis in industrial systems[M]. London: Springer, 2001: 85-98.
- [5] Yu C J, Yin S, Okyay K A, et al. Data-driven monitoring and safety control of industrial cyber-physical systems: Basics and beyond[J]. IEEE Access, 2018, 6: 47374-47384.
- [6] Yu C J, Yin S. Recent advances in key-performance-indicator oriented prognosis and diagnosis with a matlab toolbox: DB-KIT[C]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2849-2858.
- [7] Zhou D, Li G, Qin S J. Total projection to latent structures for process monitoring[J]. Aiche Journal, 2010, 56(1): 168-178.
- [8] Wold S, Kettaneh-Wold N, Skagerberg B. Non-linear PLS modeling[J]. Chemomintellabsyst, 1989, 7(1): 53-65.
- [9] Yin S, Ding S X, Zhang P, et al. Study on modifications of PLS approach for process monitoring[J]. IFAC Proceedings Volumes, 2011, 44(1): 12389-12394.
- [10] Peng K, Zhang K, You B, et al. Quality-relevant fault monitoring based on efficient projection to latent structures with application to hot strip mill process[J]. Control Theory and Applications Iet, 2015, 9(7): 1135-1145.
- [11] Wang G, Yin S. Quality-related fault detection approach based on orthogonal signal correction and modified PLS[J]. IEEE Transactions on Industrial Informatics, 2017, 11(2): 398-405.
- [12] Yin S, Wang G, Gao H. Data-driven process monitoring based on modified orthogonal projections to latent structures[J]. IEEE Transactions on Control Systems Technology, 2016, 24(4): 1480-1487.
- [13] Wold S, Antti H, Lindgren F, et al. Orthogonal signal correction of near-infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems, 1998, 44(1): 175-185.
- [14] Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS)[J]. Journal of Chemometrics, 2010, 16(3): 119-128.
- [15] Sjöblom J, Svensson O, Josefson M. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems, 1998, 44: 229-244.
- [16] Fearn T. On orthogonal signal correction[J]. Chemometrics and Intelligent Laboratory Systems, 2000, 50(1): 47-52.
- [17] Andersson C A. Direct orthogonalization[J]. Chemometrics and Intelligent Laboratory Systems, 1999, 47(1): 51-63.
- [18] Svensson O, Kourti T, Macgregor J F. An investigation of orthogonal signal correction algorithms and their characteristics[J]. Journal of Chemometrics, 2010, 16(4): 176-188.
- [19] Qin S J. Statistical process monitoring: Basics and beyond[J]. Journal of Chemometrics, 2003, 17(8/9): 480-502.
- [20] Shen Y, Ding S X, Haghani A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process[J]. Journal of Process Control, 2012, 22(9): 1567-1581.

#### 作者简介

孔祥玉(1967—), 男, 副教授, 博士生导师, 从事随机系统分析、非线性系统建模及其应用、复杂系统故障诊断等研究, E-mail: xiangyukong01@163.com;

罗家宇(1994—), 男, 硕士生, 从事数据驱动的复杂系统下故障监测与诊断的研究, E-mail: 540629964@qq.com;

张琪(1990—), 女, 副教授, 博士, 从事故障诊断、非线性滤波等研究, E-mail: 15829799698@163.com;

曹泽豪(1993—), 男, 硕士生, 从事复杂系统故障监测与诊断的研究, E-mail: 578021174@qq.com.

(责任编辑: 郑晓蕾)